

Income Taxation of Top Earners in Honduras: Linking Personal & Corporate Taxes

Code Repository Documentation

Thiago Scot*
Pierre Bachas†
Santiago Cesteros‡
Tatiana Flores§
Gabriel Oqueli¶

November 2024

*Development Impact Evaluation, World Bank. Email: tscot@worldbank.org

†Development Research Group, World Bank & EU Tax Observatory. Email: pbachas@worldbank.org

‡Fiscal Policy and Sustainable Growth Unit, World Bank. Email: scesteros@worldbank.org

§Development Impact Evaluation, World Bank. Email: rfloresibarra@worldbank.org

¶Development Impact Evaluation, World Bank. Email: gabriel.oqueli@unah.hn

Contents

1	Overview	3
2	Data Availability and Provenance Statements	3
2.1	Statement about Rights	3
2.2	Summary of Availability	3
2.3	Details on each Data Source	3
3	Datasets	4
3.1	Input datasets	4
3.2	Datasets for analysis	4
4	Computational Requirements	4
5	Instructions to replicators	6
5.1	Folder structure	6
6	List of tables, figures and programs	7
6.1	Mapping of tables and analysis code	7
6.2	Mapping of figures and analysis code	7
7	Description of programs and code	7
7.1	Creation of data sets for analysis	8
7.2	Analysis	10
7.3	Other Codes	11
7.4	Programs	11

1 Overview

The codes in this repository replicate the tables and figures from “Income Taxation of Top Earners in Honduras: Linking Personal & Corporate Taxes”, by Scot, Bachas, Cesteros, Flores, and Oqueli. The replication folder contains the codes to go from the raw administrative data to the results in the paper. This documentation is structured as follows. Section 2 describes the data sources and their availability. Section 3 describes the datasets used in the analysis. Section 4 provides details on the computational requirements. Section 5 provides instructions to replicators. Section 6 provides a mapping between the codes and the tables and figures of the paper. Finally, section 7 describes the codes, including data creation codes (section 7.1), analysis codes (section 7.2), and programs (section 7.4).

2 Data Availability and Provenance Statements

- This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

2.1 Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

2.2 Summary of Availability

- All data **are** publicly available.
- Some data **cannot be made** publicly available.
- No data can be made** publicly available.

The administrative data for this project is owned by the Honduran tax authority, the Servicio de Administración de Rentas (SAR). The data were made available to us exclusively for the purpose of this research project through collaboration agreements between the corresponding author and the government agency.

Individuals interested in accessing the data for replication purposes can contact Thiago Scot (tscot@worldbank.org). We will provide reasonable assistance to requests for clarification and replication, and will support application for data access through an agreement with SAR.

Researchers interested in obtaining the data for their own analyses can directly contact the Servicio de Administracion de Rentas de Honduras (SAR)¹.

2.3 Details on each Data Source

The analysis is based on de-identified administrative raw data provided by the Honduran tax authority, the Servicio de Administración de Rentas (SAR). The main datasets include information on rental and

¹Edgardo Espinal (eespinal@sar.gob.hn).

capital gains, gender and other personal information (e.g., marital status or location), shareholder information linked to the firms to which the individuals have any participation, corporate income tax (CIT), distributed dividends at the individual level. Data is available mainly from 2003 to 2020.

Data on [inflation](#) and [GDP deflators](#) was obtained from [DataBank](#). Data on PPP exchange rate was obtained from the World Inequality Database, using their [Stata *wid* package](#).

3 Datasets

3.1 Input datasets

Dataset	Description	Notes	Provided
'year'.csv	Corporate income tax raw data.	Confidential	No
21012023_2003_2010.csv	Corporate income tax raw data.	Confidential	No
clean_shareholder_list.dta	Raw full list of Shareholders.	Confidential	No
clean_shareholder_list_IndividualOnly	Raw full list of individual shareholders	Confidential	No
DJ_ISR.csv	Corporate income tax raw data.	Confidential	No
DMR_113_v2.csv	Distributed dividends data.	Confidential	No
ISR_PJ_v3.dta	Corporate income tax raw data.	Confidential	No
Base_Edad_Genero.csv	Gender data.	Confidential	No
Alquileres.csv	Rental income raw data.	Confidential	No
Ganacias_Capital.csv	Capital gains raw data.	Confidential	No

3.2 Datasets for analysis

Dataset	Description	Notes	Provided
inflation.dta	Inflation data.	Public	No
ISRPJ2021_withShareholderInfo.dta	Clean data with shareholders information	Confidential	No
TaxData_Analysis_'year'.dta	Clean dataset at the individual level with information tax liability and income.	Confidential	No
WID_PPP.dta	PPP exchange rate data.	Public	No

4 Computational Requirements

Software requirements

The data creation and analysis code is written in **Stata**. Version 16 is used.

Controlled randomness

Memory and runtime requirements

We suggest the use of a processor with at least 8 cores and a RAM of 16 GB. With those settings, the replicator can expect full data creation to take about 5 minutes. Around 2 minutes can be expected to run the full data analysis.

5 Instructions to replicators

5.1 Folder structure

For full replication of the project, we suggest the following, complete folder structure.

- code
- data
 - Inputs
 - AuxData
 - RawTaxData
 - Desigualdad_Jan23
 - MostUpdated_Apr23
 - Outputs
 - AnalysisTaxData
 - Auxiliary
 - ShareholdersData
- output
 - tables
 - figures

This folder structure is reflected in the do-file `0_master_file.do`, which we provide in the code repository. We provide the do-files to create the datasets used in the project and the subsequent analysis. The `0_master_file.do` do-file define *globals* in Stata, which correspond to the above folder tree. The codes described in section 7 rely on these folder paths.

6 List of tables, figures and programs

The provided code reproduces:

- ✓ All numbers provided in text in the paper
- ✓ All tables and figures in the paper
- Selected tables and figures in the paper, as explained and justified below.

6.1 Mapping of tables and analysis code

Table 1 provides a mapping between all tables of the paper (including the online appendix) and the codes producing these results. The codes generate *.tex* files containing the content of the tables.

Table 1: Mapping of tables and analysis code

Table	Code	File Name
Table 1	Effective_Tax_Rates_Other.do	Table_1.tex

6.2 Mapping of figures and analysis code

Table 2 provides a mapping between all figures of the paper (including the online appendix) and the code producing these results. The codes generate *.pdf* or *.png* files containing the graphs.

Table 2: Mapping of figures and analysis code

Figure	Panels	Code	File Name
Figure 1	A, B	Effective_Tax_Rates_top15_intervals.do	A: Figure_1_A.png B: Figure_1_B.png
Figure 2		Effective_Tax_Rates_top15_intervals.do	Figure_2.png
Figure 3	A, B	Effective_Tax_Rates_top15_intervals.do	A: Figure_3_A.png B: Figure_3_B.png
Figure 4	A, B	30b_DescriptivesPJ.do	A: Figure_4_A.png B: Figure_4_B.png
Figure 5	A, B, C	Effective_Tax_Rates_top15_intervals.do	A: Figure_5_A.png B: Figure_5_B.png C: Figure_5_C.png
Figure A1		Effective_Tax_Rates_top15_intervals.do	Figure_A1.png

7 Description of programs and code

The project is organized as follows. First, we use the raw data files to create the datasets used in the analysis. These codes are described in section 7.1. Second, we describe the data analysis code in section 7.2.

Some of the code rely on programs downloaded from the Statistical Software Components (SSC) archive. These programs are described in section 7.4.

Every code in our project folder loads a setup-script as the first line of code. This file is titled `0_master_file.do` and is part of the code repository. Global paths to the different folders of the

project (input files, output tables, logs, etc.) are established through this call. The replicator can adjust the links in these files centrally without needing to adjust the (relative) paths in the specific cleaning or analysis parts. Moreover, `0_master_file.do` installs all packages needed in the project that are downloaded from SSC.

7.1 Creation of data sets for analysis

1a_RentalCapGains.do

This do-file creates `RentalIncome.dta` and `CapitalGains.dta` datasets, which are used as intermediate datasets in the `1d_Inequality_Clean_loop` and `1e_Inequality_Clean_2003_2010` do files.

Input data:

```
$databases/Inputs/RawTaxData/MostUpdated_Apr23/Alquileres.csv
```

```
$databases/Inputs/RawTaxData/MostUpdated_Apr23/Ganacias_Capital.csv
```

Output data:

```
$databases/Output/CleanTaxData/RentalIncome.dta
```

```
$databases/Output/CleanTaxData/CapitalGains.dta
```

1b_GenderData.do

This do-file creates `Gender_age_dataset.dta` dataset, which has gender information at the individual level and is used as intermediate dataset in the `1d_Inequality_Clean_loop` and `1e_Inequality_Clean_2003_2010` do files.

Input data:

```
$databases/Inputs/RawTaxData/Desigualdad_Jan23/Base_Edad_Genero.csv
```

Output data:

```
$databases/Output/CleanTaxData/Gender_age_dataset.dta
```

1cc_CIT_Clean.do

This do-file creates `CIT_clean.dta` dataset, which has corporate income tax information and is used as intermediate dataset in the `1c_UndistributedProfits` do file.

Input data:

```
$databases/Inputs/RawTaxData/ISR_PJ_v3.dta
```

Output data:

```
$databases/Output/Auxiliary/CIT_clean.dta
```

1c_UndistributedProfits.do

This do-file creates the `UndistributedProfits_relation.dta` and the `UndistributedProfits_individual` datasets, which are used as intermediate datasets in the `1e_Inequality_Clean_2003_2010`, and the `1d_Inequality_Clean_loop` do files.

Input data:

```
$databases/Inputs/AuxData/clean_shareholder_list_IndividualOnly.dta
```

```
$databases/Inputs/RawTaxData/DMR_113_v2.csv
```

```
$databases/Output/Auxiliary/CIT_clean.dta
```

Output data:

```
$databases/Output/Auxiliary/UndistributedProfits_relation.dta
```

```
$databases/Output/Auxiliary/UndistributedProfits_individual.dta
```

1d_Inequality_Clean_loop.do

This do-file creates the TaxData_‘year’.dta datasets for the 2011-2020 period, which have inequality measures related data and are used as intermediate datasets in the 2d_Inequality_CreateVars_loop do file.

Input data:

```
$databases/Inputs/RawTaxData/MostUpdated_Apr23/‘year’.csv
```

```
$databases/Output/CleanTaxData/CapitalGains.dta
```

```
$databases/Output/CleanTaxData/RentalIncome.dta
```

```
$databases/Output/Auxiliary/UndistributedProfits_individual.dta
```

```
$databases/Output/CleanTaxData/Gender_age_dataset
```

Output data:

```
$databases/Output/CleanTaxData/TaxData_‘year’.dta
```

1e_Inequality_Clean_2003_2010.do

This do-file creates the TaxData_‘year’.dta datasets for the 2003-2010 period, which have inequality measures related data and are used as intermediate datasets in the 2d_Inequality_CreateVars_loop do file.

Input data:

```
$databases/Inputs/RawTaxData/MostUpdated_Apr23/‘year’.csv
```

```
$databases/Output/CleanTaxData/CapitalGains.dta
```

```
$databases/Output/CleanTaxData/RentalIncome.dta
```

```
$databases/Output/Auxiliary/UndistributedProfits_individual.dta
```

```
$databases/Output/CleanTaxData/Gender_age_dataset
```

Output data:

```
$databases/Output/CleanTaxData/TaxData_‘year’.dta
```

2a_Inequality_CreateVars_loop.do

This do-file creates the TaxData_Analysis_‘year’.dta datasets, which have information on income and tax liability.

Input data:

```
$databases/Output/CleanTaxData/TaxData_‘year’.dta
```

Output data:

```
$databases/Output/AnalysisTaxData/TaxData_Analysis_‘year’.dta
```

3a_PrepareForDescriptives.do

This do-file creates the ISRPJ2021_withShareholderInfo.dta dataset, which is later used to generate Figure 4.

Input data:

```
$databases/clean_shareholder_list.dta
$databases/DMR_113_v2.csv
$databases/DJ_ISR.csv
```

Output data:

```
$databases/shareholder_PJlevel.dta
$databases/ISRPJ2021_withShareholderInfo.dta
```

7.2 Analysis

Effective_Tax_Rates_top15_intervals.do

This do-file creates Figure 1, Figure 2, Figure 3, Figure 5, and Figure A1. Figure 1, Panel (a) presents the (log) average income for each bin of the distribution in local currency (Lempiras). Figure 1, Panel (b) presents the composition of total income for each bin of the distribution. Figure 2 displays the average effective tax rate (ETR) for each bin of the distribution. Figure 3, Panel (a) plots the average effective tax rate (ETR) for each bin of the distribution for three different taxes, while Panel (b) displays the median effective tax rate (ETR) for each bin of the distribution together with the percentiles 25 and 75. Figure 5, Panel (a) displays the share of individuals in each bin that are shareholders in any corporation and those that are shareholders in top decile corporations. Panel (b) displays the mean and median number of firms at the bin level in which individuals are shareholders. Panel (c) displays the mean share of shareholders that have shareholdings in at least one firm with exempted income and the mean share of exempted profits at the bin level. Finally, Figure A1 displays the effective tax rate (ETR) for each bin of the distribution for four different years (2013, 2016, 2018, and 2019).

Input files:

```
$databases/Output/AnalysisTaxData/TaxData_Analysis_‘year’.dta
$databases/Inputs/AuxData/inflation.dta
$databases/Inputs/AuxData/WID_PPP.dta
```

Output files:

```
$figures/Figure_1_A.png
$databases/Figure_1_B.png
$databases/Figure_2.png
$databases/Figure_3_A.png
$databases/Figure_3_B.png
$databases/Figure_5_A.png
$databases/Figure_5_B.png
$databases/Figure_5_C.png
```

\$figures/Figure_A1.png

Effective_Tax_Rates_Other.do

This do-file creates Table 1, which describes the income distribution for the top 1.5% of the distribution.

Input files:

\$databases/Output/AnalysisTaxData/TaxData_Analysis_‘year’.dta

\$databases/Inputs/AuxData/inflation.dta

\$databases/Inputs/AuxData/WID_PPP.dta

Output files:

\$tables/Table_1.tex

3b_DescriptivesPJ.do

This do-file creates Figure 4, Panel (a) presents the average share of firms with at least one foreign shareholder and the average share of firms with at least one shareholder who is a legal person for each company group. Panel (b) shows the average share of revenue and the average share of profits that is attributable to a shareholder with a valid RTN for each company group.

Input files:

\$databases/Output/ISRPJ2021_withShareholderInfo.dta

Output files:

\$figures/Figure_4_A.png

\$figures/Figure_4_B.png

7.3 Other Codes

IncomeTaxBrackets.do

This do-file provides locals with information on the minimum wage and the thresholds (tax brackets) for each category of the progressive income tax scheme for each year.

7.4 Programs

Programs installed via SSC

Our code utilizes the following commands installed via SSC manually in `0_master_file.do`:

- **egenmore:**

Nicholas J. Cox, 2000. “EGENMORE: Stata modules to extend the generate function,” Statistical Software Components S386401, Boston College Department of Economics.

- **gtools:**

Mauricio Caceres Bravo, 2018. “GTOOLS: Stata module to provide a fast implementation of common group commands,” Statistical Software Components S458514, Boston College Department of Economics, revised 03 Apr 2019.

- **wid:**

Blanchet, T. (2021). "WID: Stata module to download data from the World Inequality Database (WID. world)".
